

PAPER • OPEN ACCESS

Developing competency profiles of IT specialists based on semantic analysis of vacancies

To cite this article: Aleksey Gintciak *et al* 2019 *J. Phys.: Conf. Ser.* **1405** 012003

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

Developing competency profiles of IT specialists based on semantic analysis of vacancies

Aleksey Gintciak¹, Andrey Rytik², Denis Terin², Marina Kudaka³, Dmitriy Kaisyn⁴, Aleksey Ignatev⁵, Aleksey Marakhtanov⁶ and Oksana Gladkih⁷

¹ Peter the Great St. Petersburg Polytechnic University, 195251, St. Petersburg, Russia

² Saratov State University, 410012, Saratov, Russia

³ Cherepovets State University, 162600, Cherepovets, Russia

⁴ Moscow Aviation Institute, 125080, Moscow, Russia

⁵ Yaroslavl State Technical University, 150023, Yaroslavl, Russia

⁶ Petrozavodsk State University, 185910, Petrozavodsk, Russia

⁷ MBA-Center of Ural Federal University, 620002, Ekaterinburg, Russia

E-mail: kudakam@mail.ru

Abstract. The article focuses on applying Big Data processing methods and tools for creating specialists' competency profiles. It covers the approach developed to formulate the key competencies required for the IT vacancies. The first stage of the research includes carrying out a semantic analysis of the IT vacancies database. The second stage includes clustering of the results of semantic analysis and using them for building the map of the competencies of IT specialists. The modularity score reveals a significant isolation of the specialties in this area. The performed approach can be implemented in designing the professional standards and competency based education programs.

1. Introduction

Currently, preparing labour resources that match the needs of demand on the labor market is of strategic importance for the integral development of various industries. In modern conditions, one of the most promising and dynamically developing industries is the IT sector, where much attention is paid to raising qualified personnel with required professional competencies. Today this labor market is critically understaffed and there is an acute shortage of graduates with the necessary competencies. Therefore, for effective training of IT personnel, it is very important to develop profiles of specialists' competencies which define the requirements of the level of knowledge, skills and personal qualities of a specialist. In essence, a competency profile is a set of key specialist skills needed by an organization for its effective development. At the same time, it is important to understand that for various industries it is appropriate to develop various competency profiles, which would determine the unique characteristics of the activity.

The need for developing competency profiles containing the least possible set of key competencies is also determined by the fact that companies operating in the same industry and adhering to similar business models often have different sets of formal requirements when recruiting employees of the same functional role. Competency mapping is an effective tool for solving this problem and unify the formalization of requirements for employees.



The expert method for constructing dynamic maps of key competencies is limited and inefficient. A prime example of redundancy, on the one hand, and the limitations of the expert method, on the other hand, is illustrated in [1], which describes the competency model of a specialist in the field of computer security. Another example is a skill map and a model for the development of team leaders [2].

Under the comprehensive transition to a digital economy, the question is not only in determining who, how, what and how much to learn, but also in managing the a transformational funnel that arises when tasks change too quickly and it becomes difficult to predict the demand of enterprises for specific competencies [3], and sometimes it's completely impossible [4].

In our opinion, one of the effective ways to overcome these identified organizational contradictions is to develop digital profiles of specialists and to use the basic tools of processing large amounts of data in combination with semantic analysis methods. Functionality sufficient to process clustering the outcome of the semantic analysis of vacancies is provided by such application software as Gephi [5] or Tableau [6].

2. Approaches to the multidimensional data analysis

The process of building a specialist's digital profile, as well as the analysis of job databases, is naturally associated with the creation of a machine algorithm for processing text and number datasets. Measurement of several parameters (object properties) at once in one experiment determines the multidimensionality of statistical analysis and the wide field for its application [7]. The multidimensional statistical analysis includes: factor analysis; discriminant analysis; cluster analysis; multidimensional scaling; quality control techniques [8]. Particularly cluster analysis is successfully applied for solving the problem of creating a profile of competencies: this method allows to divide the studied set of objects into groups of "similar" or interdependent objects called clusters and identify a cluster, for example, of the most popular and highly paid specializations. Using classification methods, we assume that certain classes of classified objects correspond with the desired classes, which can be considered as points of some attribute space, etc. Choosing a model of this kind essentially means choosing a mathematical method. Classification methods, combining similar objects into clusters, "compress" the data matrix row by row.

3. Building a competency map

Implementing BigData instruments helps to obtain more informed conclusions comparing to the use of only conventional methods of data processing, i.e. statistical methods. This research was carried out using cluster and semantic analysis. The result is performed by a map of the specialists' competencies regarding the field of information technologies. Similar maps were created before, but they were based solely on expert opinion and that is why were more subjective.

In order to build a competency map for IT specialists, it is necessary to solve a number of sequential tasks: selecting source data, compilation and creating a database, analyzing a database, database cleansing, searching for interdependencies from all data to obtain statistically significant conclusions, cause-effect analysis and forecasting the development of the situation on the basis of processed and structured data.

3.1. Source Dataset

In our research we use a test database of vacancies derived from the portal "Work in Russia" [9]. The base includes more than three million vacancies actual for the period from October 2015 to June 2017 performed by Employment Centers, personal accounts of employers on the portal, information systems of employers. The database is a set of csv files, divided by the professional profile and region. Detailed information is available for each vacancy, including the date of publication of the vacancy, position, proposed salary, and (the most important for this study) a set of requirements for a candidate presented by the employer. This set of requirements within the research is identified with a list of specialists' competencies of required in the whole IT-industry.

The created dataset consists of vacancies from the sections of the source database named "Information Technologies", "Information Technologies, Telecommunications, Communication" of all regions of the Russian Federation. The size of this sample was 26,450 rows. The dataset includes only the fields relating to the subject of the research: requirements, qualifications, date of addition, salary.

3.2. Data preparation (wrangling)

In order to prepare the dataset for further work we used spreadsheet applications which allow to edit csv files and removed the records containing the following elements:

- html tags (<p>,
, , etc.);
- duplicates;
- typos;
- vacancies with missing qualifications;
- entry errors (the text in the qualification field that is not related to requirements: "according to the job description", "lack of criminal record", "except days off", etc.);
- incorrect indication of upper and lower case.

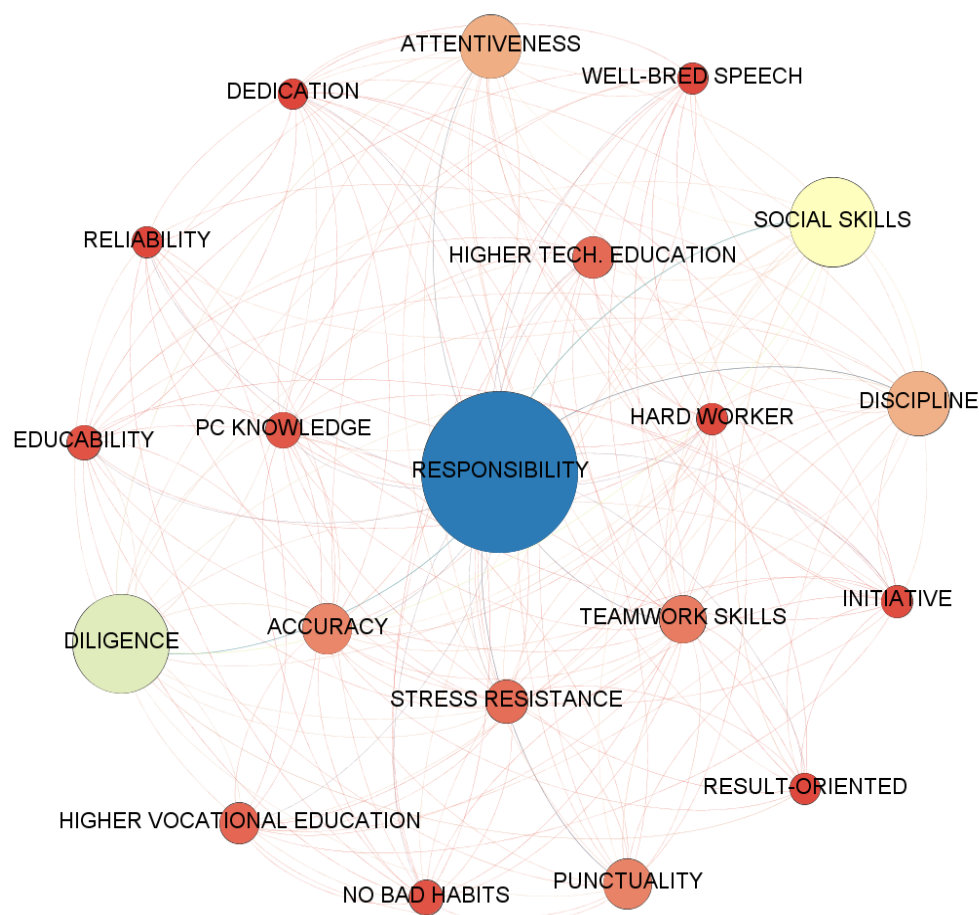


Figure 1. Graph of general and social competencies

3.3. Visualization of the result

After completing the processes of preparing and cleaning the dataset it was analyzed and visualized in the form of a graph by weaving the terms of the "qualification" field. The repeating phrases in these fields were defined as the vertices of the graph, and the relationships between the phrases (how often the same requirements are found within the same vacancy) are defined as edges of the graph.

As the result there was generated a weighted graph of 20,257 vertices and 106,985 edges. The average degree of each vertex is 10.5. The graph using the Louvain Method [10] identified 2956 clusters containing interconnected vertices of competencies and vacancies. The frequency is determined as the weight of the edge – the number of vacancies which included both phrases (vertices) at once. The weight of the vertices – the frequency of occurrence of the phrase in the dataset.

Initial visual analysis of the resulting graph showed that the highest frequency of occurrence was performed by soft skills, such as "responsibility", "sociability", "diligence", "teamwork", etc. Within the dataset, general and professional requirements are not divided into different fields. For research purposes, competencies had to be divided into professional (hard skills) and soft skills manually. The graph of competencies not directly related to IT is presented in Figure 1.

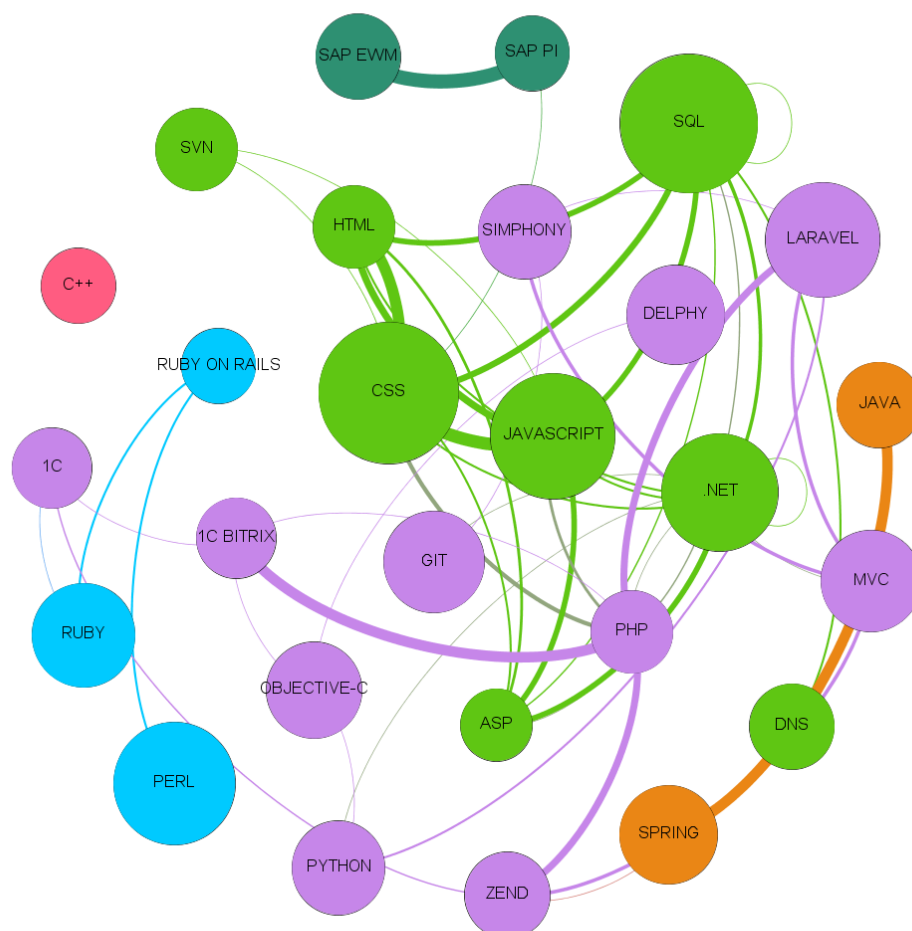


Figure 2. General graph of highly specialized professional competencies

For research purposes, the IT-related competencies are of most interest, and therefore the general and social competencies were removed from the general graph. For this, a soft skills dictionary [11]

was compiled, and the vertices corresponding to the values in the dictionary were marked as soft skills. This allowed view and analyze separately graphs with soft skills and highly specialized professional (Figure 2) competencies.

The graph of highly specialized competencies as a result of the above mentioned processes contains the key requirements for IT specialists.

3.4. Modularity

The next step is to examine how sustainably the algorithm allocates clusters depending on the degree of the graph structure. The importance of the correct clusterization is determined by the fact that significant conclusions are often based on the revealed relationships between the elements of the subsets.

For evaluating the correctness of cluster allocation the modularity score is used [12]. Using this measure, one can evaluate the quality of the graph structure. The methodology for calculating modularity for weighted networks is given in [13].

Analytically, modularity is equivalent to the fraction of edges that connect vertices in the subset i , minus the expected fraction of edges in the network with a similar division into subsets, but with random connections between the vertices.

Modularity is a scalar quantity and can take values in the range $[-1, 1]$. As a rule, usually the measure falls in the range from 0.3 to 0.7. This range indicates an acceptable allocation to subsets. Higher values are rare and may indicate degeneracy of the sample. If the modularity value is 0, then the relations between the subsets can be considered random [14]

The problem of optimizing modularity is quite complicated for calculations and can be solved in different ways. We used the Louvain Method [10], the calculation was fulfilled using weighted modularity, the formula is given above [13]. The method was chosen due to high speed and accuracy of calculations. The tests carried out confirmed that the algorithm used allows to divide the network into subsets with modularity close to optimal [15].

The modularity score calculated from the generated graph is 0.895. A quantitative assessment of the quality of dividing a network into subsets is considered quite high. This indicates a qualitative distribution in the graph presented, the relationship between the competencies of IT specialists can be considered statistically significant. The interconnections between the vertices within the communities are strong and dense, while the interconnections between the vertices of different communities are rare.

At the same time, the number of subsets is 2956. The Louvain Method algorithm allows to recalculate modularity with a different resolution parameter. Recalculation even with a high resolution does not significantly reduce the number of subsets. A similar result is predictable, since the database used for research purposes additionally needs a thorough cleaning. However, even analyzing partially contaminated data, the conclusions of the research have sufficient practical significance.

4. Interpretation of the result

Regarding the high modularity index of the competency graph of IT specialists, provided a high degree of reliability of the source data, we can conclude that specialties in the IT sphere have quite significant differences in the set of required competencies. In other words, currently there is no such generalized concept as an IT specialist, but in this area there is a certain number of highly specialized vacancies, each of them is characterized by its own set of required competencies. For example, a distinguished multicluster is defined as characterizing the specialty of a web developer (Figure 3). It's key competencies are knowledge of programming languages such as JavaScript, HTML, CSS, SQL, PHP, ASP, .NET. This set of programming languages, closely identified with competencies, corresponds to the generally accepted ideas about web development in the world community at the moment, which further confirms the adequacy of the results obtained in the study. It should be noted that this cluster, when examined in more detail, can be divided into more highly specialized areas, for example, typesetting, front-end and back-end development, and working with databases.

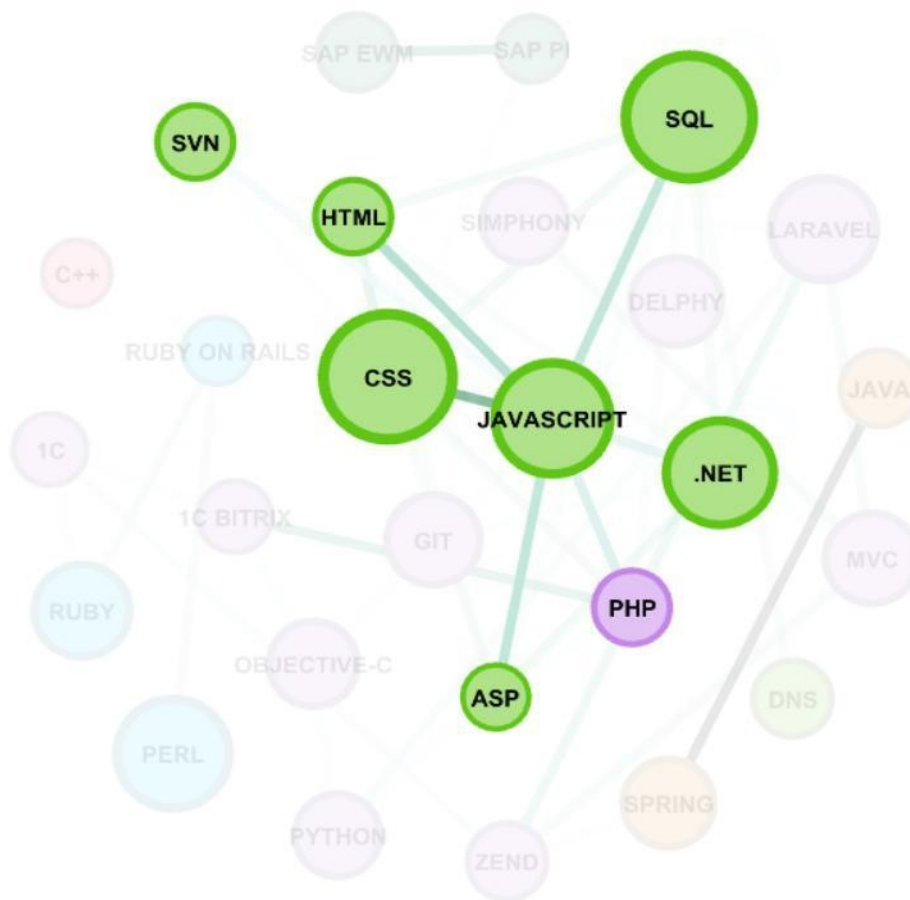


Figure 3. Web developer competency map

In this way, we can evaluate the specialties of the developer of products accompanying SAP or other corporate ERP systems; system administrator; network administrator; automated or manual testing engineers; other developers. Moreover, the vertices of the graph that have the greatest connectivity with the neighboring ones (in Figure 3 correspond to the volume of the vertex) are the highest priority for the considered specialty.

Thus, using the methods and tools presented in this study makes it possible to build full-scale competency maps for IT specialties. The outcome can be used in the development of professional standards and educational programs (both primary and secondary education).

5. Conclusion

As part of this research, a semantic analysis of the vacancies database of IT specialists was carried out regarding the required qualifications, general professional and highly specialized competencies; the outcome of the semantic analysis was processed using clustering techniques and on this basis the result was suggested as an approach for creating competency maps of specialists in the related field.

The proposed method also can be used for building competency maps of specialists in other professional fields, adjusted to the their specifics.

The results of the analytical processing of documents on monitoring subject areas are used in mathematical forecasting and analysis of markets and social situations. It can be concluded that the development of IT specialists' competency profiles based on the semantic analysis of vacancies allows us to evaluate not only the real demand for employees in IT industry, but also to predict it.

A promising area of research is studying the resumes in the relevant industry for building competency maps of the specialists currently present in the labor market. Assuming these maps to indicate the labor market supply, and assuming competency maps similar to the map built in this research to indicate the demand on the labour market, we can verify whether supply and demand match or there can be revealed quantitative or qualitative gaps between them. Such further investigation will contribute greatly to the study of the current state of labor relations in the related field and will help to map out the steps to reduce the gap at the macroscopic level.

Another area of investigation of great practical importance is the study of maps of the required competencies in the context of time. Considering the sequence of maps created in different time intervals makes it possible to track trends of demand for individual competencies or entire specialties, which will lead to the possibility of forecasting demand for specific specialists in the near future and, as a result, outstripping the training of these specialists under the concept of competency based education.

References

- [1] Gridin A *Competence model of a specialist in computer security* [Online] <https://habr.com/ru/post/182176/>
- [2] Tsyganov S et al *Teamlead roadmap* [Online] <https://github.com/tlbootcamp/tlroadmap>
- [3] Kudryavtseva Y *How staff services respond to the challenges of the digital economy* [Online] <https://www.vedomosti.ru/management/blogs/2019/04/01/797980-sluzhbam-personala>
- [4] Bolsunovskaya M V, Gintciak A M, Yudina I V and Kozlovskii P S Comparative analysis of approaches to the specialists' competences development in innovative industries *Advances in Social Science, Education and Humanities Research* **298** <https://doi.org/10.2991/essd-19.2019.18>.
- [5] Software Gephi official website *Gephi.org* [Online] <https://gephi.org>
- [6] Software Tableau official website 2003–2019 *Tableau Software* [Online] <https://www.tableau.com>
- [7] Chapaikina N E 2012 Semantic analysis of texts. Basic Features *Molodoy uchoniyy* **5** pp 112–115
- [8] Dezin A A and Aleksanova I 2018 *Multidimensional analysis and discrete models* (Book. Steklov Mathematical Institute, Russian Academy of Sciences, Moscow)
- [9] The official resource of the Federal Service for Labor and Employment
- [10] Vincent D Blondel and Jean-Loup Guillaume 2008 Renaud Lambiotte and Etienne Lefebvre Fast unfolding of communities in large networks *Journal of Statistical Mechanics: Theory and Experiment* (JSTAT) P10008 <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- [11] Klochkova E S, Bolsunovskaya M V and Shirokova SV 2018 The significance of humanities in engineering education *2018 XVII Russian Scientific and Practical Conference on Planning and Teaching Engineering Staff for the Industrial and Economic Complex of the Region* (PTES) DOI: 10.1109/PTES.2018.8604199
- [12] Newman M E J 2006 Modularity and community structure in networks *Proceedings of the National Academy of Sciences* **103**(23):8577–8582 <https://doi.org/10.1073/pnas.0601602103>
- [13] Girvan M and Newman M E J 2002. Community structure in social and biological networks *Proceedings of the National Academy of Sciences* **99**(12):7821–7826 <https://doi.org/10.1073/pnas.122653799>
- [14] Girvan M and Newman M E J 2004 Finding and evaluating community structure in networks *Phys. Rev. E* **69** 026113 <https://doi.org/10.1103/PhysRevE.69.026113>
- [15] An electronic resource with a detailed description of the functionality and implemented algorithms of Gephi [Online] <https://github.com/gephi/gephi/wiki/Modularity> (accessed 07/08/2019)